# GENES V

Benjamin Lewin

OXFORD UNIVERSITY PRESS
Oxford New York Tokyo
1994

It is possible to assign a specific end to every intron by relying on the homology of exon–intron junctions. They can all be aligned to conform to the consensus sequence given in **Figure 31.2**. (In this as in other cases, we write just the sequence of the DNA strand that is identical with the RNA product.)

The subscripts indicate the percent occurrence of the specified base (or type of base) at each consensus position. The really high conservation is found only *immediately within the intron* at the presumed junctions. This identifies the sequence of a generic intron as:

GT. . . . .AG

Because the intron defined in this way starts with the dinucleotide GT and ends with the dinucleotide AG, the junctions are often described as conforming to the **GT-AG rule**.

Note that the two sites have different sequences and so they define the ends of the intron *directionally*. They are named proceeding from left to right along the intron, that is, as the **left (or 5′)** and **right (or 3′)** splicing sites. Sometimes they are called the **donor** and **acceptor** sites. The consensus sequences are implicated as the sites recognized in splicing by point mutations that prevent splicing *in vivo* and *in vitro*.

The GT-AG rule describes the splicing sites of nuclear genes of many (perhaps all) eukaryotes. This implies that there is a common mechanism for splicing the introns out of RNA. The consensus does not apply to the introns of mitochondria and chloroplasts, nor to the yeast tRNA genes. (We discuss these situations later.)

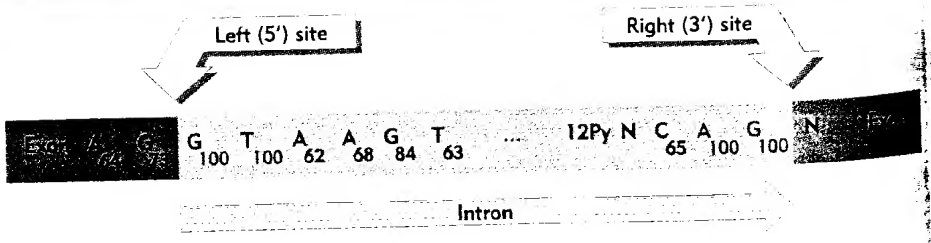
A typical mammalian mRNA has many introns. What ensures that the correct pairs of junctions are spliced together? We can imagine two types of principle that might be responsible for pairing the appropriate 5′ and 3′ sites:

- It could be an *intrinsic property* of the RNA to connect the sites at the ends of a particular intron, for example, because of base pairing involving these regions.
- Or all 5′ sites may be functionally equivalent and all 3′ sites may be similarly indistinguishable, but splicing could follow *rules* that ensure a 5′ site is always connected to the 3′ site that comes next in the RNA.

The splicing sites of nuclear RNAs do not themselves have any sequence complementarity, and nor do the surrounding sequences offer the ability to pair. Experiments to construct hybrid RNA precursors show that any 5′ splicing site can in principle be connected to any 3′ splicing site. For example, when the first exon of the early SV40 transcription unit is linked to the third exon of mouse β-globin, the hybrid intron can be spliced out to generate a perfect connection between the SV40 exon and the β-globin exon. (Indeed, this capacity is the basis for the exon trapping technique reviewed previously in Figure 23.18.) Such experiments make two general points:

**Figure 31.2**

The ends of the intron are defined by the GT-AG rule.



Splicing sites are generic: they do not have specificity for individual RNA precursors, and individual precursors do not convey specific information (such as secondary structure) that is needed for splicing.

And the apparatus for splicing is not tissue specific: an RNA can usually be properly spliced by any cell, whether or not it is usually synthesized in that cell. (We discuss some exceptions in which there are tissue-specific alternative splicing patterns later.)

Here is a paradox. Probably all 5′ splicing sites look similar to the splicing apparatus, and all 3′ splicing sites look similar to it. *In principle any 5′ splicing site may be able to react with any 3′ splicing site.* But in the usual circumstances splicing occurs only between the 5′ and 3′ sites of the *same* intron. *What rules ensure that recognition of splicing sites is restricted so that only the 5′ and 3′ sites of the same intron are spliced?*

Are introns removed in a specific *order* from a particular RNA? Using RNA blotting, we can identify nuclear RNAs that represent intermediates from which some introns have been removed. Figure 31.3 shows a blot of the precursors to ovomucoid mRNA. There is a discrete series of bands, which suggests that splicing occurs via definite pathways. (If the seven introns were removed in an entirely random order, there would be more than >300 precursors with different combinations of introns, and we should not see discrete bands.)

There does not seem to be an *obligatory* pathway, since intermediates can be found in which different combinations of introns have been removed. However, there is evidence for a *preferred* pathway or pathways. When only one intron has been lost, it is virtually always 5 or 6, but either can be lost first. When two introns have been lost, 5 and 6 are again the most frequent, but there are other combinations. Intron 3 is never or very rarely lost at one of the first three splicing steps. From this pattern, we see that there is a preferred pathway in which introns are removed in the order 5/6, 7/4, 2, 3/1. But clearly there are other pathways, since (for example), there are some molecules in which 4 or 7 is lost last. A caveat

**Figure 31.3**

Northern blotting of nuclear RNA with an ovomucoid probe identifies discrete precursors to mRNA. The contents of the more prominent bands are indicated. Photograph kindly provided by Bert O'Malley.